

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2003, 13:05:41 ; Search time 45 Seconds

(without alignments)
3814.161 Million cell updates/sec

Title: US-10-027-000-2

Perfect score: 4391
Sequence: 1 MADIDVEAIILKILIAEKVD.....DGVALRGKFTVGETYMWGV 833

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protist:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1996.5	45.5	832	3	09P6J6
2	1703	38.8	862	3	08C3Z9
3	1595.5	36.3	861	3	09HGK2
4	1143.5	26.0	772	2	066050
5	1095.5	24.9	754	2	008331
6	982	22.4	720	2	060038
7	959	21.8	721	16	09WKN2
8	958.5	21.8	828	2	051843
9	957	21.8	859	16	087852
10	954	21.7	808	2	070021
11	946.5	21.6	762	16	09A7G7
12	944.5	21.5	809	2	09ZGH5
13	941	21.4	781	2	09F3Y0
14	938.5	21.4	793	16	08ZCZ9
15	919.5	20.9	732	2	09RH03
16	874.5	19.9	752	2	P96316

17	874	19.9	758	16	09A9L2	09A912 caulobacter
18	860	19.6	735	2	024749	024749 acetobacter
19	844.5	19.2	759	2	068843	068843 streptomyces
20	839.5	19.1	739	2	08RTV7	08RTV7 acetobacter
21	839	19.1	691	16	007430	007430 mycobacteri
22	838.5	19.1	926	16	09KBL8	09KBL8 bacillus ha
23	822	18.7	789	2	059716	059716 prevotella
24	818.5	18.6	784	2	09JRK1	09JRK1 thermoaer
25	806	18.4	730	2	P96090	P96090 thermoaer
26	796.5	18.1	803	3	099324	099324 septoria ly
27	776.5	17.7	754	17	097YU4	097YU4 sulfobus
28	756.5	17.2	778	2	056322	056322 thermotoga
29	756.5	17.2	821	2	09KMY5	09KMY5 thermotoga
30	747	17.0	778	16	09WXT1	09WXT1 thermotoga
31	746	17.0	823	3	074203	074203 phanerocha
32	746	17.0	823	3	09URP5	09URP5 phanerocha
33	735	16.7	802	3	08TGC6	08TGC6 phanerocha
34	731.5	16.7	744	3	012715	012715 trichoderma
35	716	16.3	764	2	013356	013356 bacteroides
36	695.5	15.8	870	3	013385	013385 cochlidiobu
37	694.5	15.8	870	3	013385	013385 cochlidiobu
38	682.5	15.5	882	16	09PEF3	09PEF3 xyliella fas
39	681	15.5	793	3	000903	000903 gaemannomy
40	675.5	15.4	986	2	059506	059506 microbispor
41	674	15.3	780	3	09UVJ6	09UVJ6 botrytis ci
42	644	14.7	896	3	09C2C4	09C2C4 neurospora
43	642.5	14.6	765	16	08Z5B5	08Z5B5 salmonella
44	633	14.4	770	2	093RB6	093RB6 streptomyces
45	628.5	14.3	765	16	08X668	08X668 escherichia

ALIGNMENTS

RESULT 1	Q9P6J6	PRELIMINARY;	PRT;	832 AA.
ID	Q9P6J6			
AC	Q9P6J6			
DT	01-OCF-2000 (TREMBLrel. 15, Created)			
DT	01-OCF-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)			
DE	Beta-glucosidase (EC 3.2.1.21), glycosyl hydrolase family 3.			
DE	SPBCL683.04.			
GN	Schizosaccharomyces pombe (Fission Yeast).			
OS	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;			
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;			
OC	Schizosaccharomycetes.			
OX	NCBI_TaxID=4896;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	McComble W.R.;			
RT	"Sequence analysis of a region of the fission yeast genome."			
RL	Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AL355920; CAB9166.1; -			
DR	InterPro; IPR002772; GH_3C.			
DR	InterPro; IPR001764; GH_3N.			
DR	Pfam; PF00933; Glyco_hydro_3; 1.			
DR	Pfam; PF01915; Glyco_hydro_3_C; 1.			
DR	PRINTS; PR00133; GLYDRLASE3.			
DR	PROSITE; PS00775; GLYCOSYL_HYDROL_F3; UNKNOWN_1.			
KW	Hydrolase.			
SO	SEQUENCE 832 AA; 93169 MW; 24C8EF2B248B02F CRC64;			
QY	Query Match	45.5%;	Score 1996.5;	DB 3; Length 832;
QY	Best Local Similarity	46.5%;	Pred. No. 6.5e-125;	
QY	Matches 389;	Conservative 157;	Mismatches 282;	Indels 9; Gaps 7;
Db	1 MADIDVEAIILKILIAEKVDLGLIDFWHRKAPKGVPLRTDGPNGVGRKFNQGV 60			
QY	1 MEHDVEAIILKILIAEKVDLGLIDFWHRKAPKGVPLRTDGPNGVGRKFNQGV 60			
QY	61 AACFPGCTSLGTFNQLTLEAGKMKGEAKIAKSAHVILGLPTINMORSPILGSGFESIG 120			

RESULT 3

09HGX2

ID 09HGX2

PRELIMINARY:

PRT: 861 AA.

AC 09HGX2;

DT 01-MAR-2001 (Tremblrel. 16, Created)

DT 01-MAR-2001 (Tremblrel. 16, last sequence update)

DT 01-MAR-2002 (Tremblrel. 20, last annotation update)

DE Putative beta glucosidase (EC 3.2.1.21).

GN BGL.

OS Agaricus bisporus (Common mushroom).

OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;

OC Agaricales; Agaricaceae; Agaricus.

OX NCBI_TaxID=5341;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=D649;

RA Morales-Almora P., Thurston C.F.;

RT "Molecular analysis of the cellulolytic genes in Agaricus bisporus";

RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AJ293760; CAC03462.1;

DR InterPro: IPR002772; GH_3C.

DR Pfam: PF00933; Glyco_hydro_3; 1.

DR Pfam: PF01915; Glyco_hydro_3; C; 1.

DR PRINTS: PR00133; GLHYDRLASE3.

KW Glycosidase; Hydrolase.

SQ SEQUENCE 861 AA: 94353 MW: 8852C0A18DB39AB3 CRC64;

Query Match 36.3%; Score 1595.5; DB 3; Length 861;

Best local similarity 42.0%; Pred. No. 4, 6e-98;

Matches 365; Conservative 147; Mismatches 306; Indels 51; Gaps 22;

QY 2 ADDIVEAIKKLTAEVNDLAGIDFHTKALPKHGVPSLFTDGPNGVGTFFNGVPA 61
 DB 7 ANAIDIKIVALTDEADILLISAGVGFHTAIEQLIPAKVSDGPRGKGNHFMGTPTA 66
 QY 62 ACPFGCTSLGTFNQTLLIEAG-KMMGKEAIKSAHVILGPTINMORSLGSGFSIGE 120
 DB 67 KCLPSSMTAKATPRDLLEEVGLKILAREKLKLSASVITLPTONIQNPLGSGFSFSE 126
 QY 121 DPLAGLGAALALRIGISTGVQATIKHFLCNDQEDRRMVOISIVTERALREIYALPQIA 180
 DB 127 DPLSGMLCSAYINGIKGIGATIKHFGVNDKEDRTGYDSI-HYTTSSGDLSPSVHA 185
 QY 181 VRDSQP--GAFMTAYNGINVSCESENKYLDMKLRKMGMDGLMSWGTYTTEVA 238
 DB 186 GSRMPLHGAIMTAYNVNNGIHVCENPELLQKILRDEMKSVMIMSDWFGTYSVDVGLNA 245
 QY 239 GLDLEMPGPPFRFG-ETLKFNVSNGKPFIVHIDORAREVLOFVKCAASGVTG---NGP 293
 DB 246 GLDLEMPGLNKRSLSEVNSIOSRKYTAKKVERAKVLELYKC-AQAPPELLDGDGL 304
 QY 294 ETVVNNPTETALLRKVNGEIVLLKNNVPLSKK--KTLIVGNKQATYHGGGSA 351
 DB 305 ERTL-DSDEDALMRKFAESIYLLKNDKRYLLPLDRKSLKVAIVGNNAKAALVSGGSA 363
 QY 352 ALRKYAVVTPEDGSKOL-ETPP-----SYTVGATTT-VPPILGEOCLTPGAGM--RWR 403
 DB 364 ALKSTYVARPAGLVKALKEVNPETVYSEGAAAKTLPLDPLETSEGQKGTGSMH 423
 QY 404 VFNEP-----GTPNRQIHDELFTKTDMLNLYDYHPRKADTYADMEGTYADE-DCY 456
 DB 424 AHEMDDSLVALPTPIKQYVDE-----TRMFISSSPAGITTKKWTLMKRYLAPREMDCK 478
 QY 457 YEGLVVGTAAYVDOLVVDNAKQVPGDAFGSATREBTGRINLVKNTYKFKTEFG 516
 DB 479 FEFILTAAGRAKLFVDKLVIDNMTRORGVEFGCGSEETGVVELKAGVKEHIYVDFC 538
 QY 517 S--APTTLKGDIVPGHSGLRGVGCVVIDQAEIEKSVLAKHDDVYIICAGINAMWET 574
 DB 539 NVKGPADDEFTETIMDSNPGIRLGGAEVEDELLNSAVSIADKADAVIIVGLNGDMET 598

QY 575 EGADRASMKLEGVLDOLLAADVAANPNTVYVQGTPEEMRMILATPAVIOATYGNBTG 634
 DB 599 EGNDDTLTALPGRIDELVQKVAANRAVVYQASGASATIMVDDVSSLTRITLGNATG 658
 QY 635 NSIADVGVGVNPSGKLSLSPKRLQDNPAFLNFTAGRTLYGEDVYGVRYFEADKD 694
 DB 659 DAIAVYIGNNHPSGKLSLTPFKRLLEDVPAHGFYSEKGYTAGGLVYKHHOHRNIE 718
 QY 695 -VNEPFGHGLSTYTFAFNSLVSHK-----DGKLSVSLSVKNTSGVPAQVQALYKPL 747
 DB 719 PLFATGHGLSTYTFNINSLRLSOPLYSGDQPDCLKATVLT--NIGNITGSOYQVYLILP 776
 QY 748 QAAKINRPYKELKGRVAV-ELDPGETKAVTIEDEKRYAAVFEEDRDMCVKICKDYEV 806
 DB 777 KTSLETHPRWQVRGFKKMDVYKPGESREVEL-VMRLSVSYWDKE---WVENGAAYDVRV 832
 QY 807 SDSS--AAKQVYALGKFLYGETYMSGV 833
 DB 833 APTSEGVGEGELLGRFVNEKFGWRGL 861

RESULT 4

066050

ID 066050

PRELIMINARY:

PRT: 772 AA.

AC 066050;

DT 01-AUG-1998 (Tremblrel. 07, Created)

DT 01-AUG-1998 (Tremblrel. 07, last sequence update)

DT 01-MAR-2002 (Tremblrel. 20, last annotation update)

DE Beta-glucosidase.

GN GLUA.

OS Ruminococcus albus.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;

OC Clostridiales; Lachnospiraceae; Ruminococcus.

OX NCBI_TaxID=1264;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=7;

RA Cho K.K.;

RT "Sequence and structure of Ruminococcus albus 7.";

RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL: U92808; AAC05445.1;

DR InterPro: IPR002772; GH_3C.

DR InterPro: IPR001764; GH_3N.

DR Pfam: PF00933; Glyco_hydro_3; 1.

DR Pfam: PF01915; Glyco_hydro_3; C; 1.

DR PRINTS: PR00133; GLHYDRLASE3.

SQ SEQUENCE 772 AA: 86092 MW: 08CDB8BEF923CC3 CRC64;

Query Match 26.0%; Score 1143.5; DB 2; Length 772;

Best local similarity 31.6%; Pred. No. 6, 7e-68;

Matches 261; Conservative 134; Mismatches 243; Indels 187; Gaps 14;

QY 4 IVEALIKKLTAEKVNDLAGIDFHTKALPKHGVPSLFTDGPNGVNR---GTKEFN-G 58
 DB 13 MDIAHMEITLLEKKSLSGADFWHTKAIERLDIPQIVSGPHLRKNDVGSNDPNA 72
 QY 59 VPACFPCTSLGTFNQTLLIEAGKMKKEAIAKSAHVILGPTINMORSLGSGFSIGE 118
 DB 73 IEAVCFPTAALACSVDRELKIDGALGEGQSEKVSILPBGCKIRKSLPCGRNFEYF 132
 QY 119 GEDPFLAGLGAALALRIGISTGVQATIKHFLCNDQEDRRMVOISIVTERALREIYALPQ 178
 DB 133 SEDPYASOMALSHIKGVOSKAGTSLKHFAANNQHRMSVSAEIDETLHEIYLAPE 192
 QY 179 IAVRDSOPGAFMTAYNGINVSCESENKPYLDGMLREMGMDGLMSDWTGYSTTEAVYA 238
 DB 193 SVTEKAPMTWKSSTYKINGEISSQKSLTTLTRKMKFQDGLVSDMKAVVDRKYGITA 252
 QY 239 GLDLEMPGPPFRFGETLKFNVSNGKPFIVHIDORAREVLOFVKCAASGVTGNGPETTVN 298
 DB 253 GLDLEMPGSKCKNDKMLAIVEDGKLSVEALDKCVRIELIDKSLDEC-----RTEM 305
 QY 299 NPET--AALLRKVNGEIVLLKNNVPLSKKKTLLIYGNKAKQATYHGGSAALRAY 357

```

Db 306 WDKERHQAQKAERSAVLLKNDHILPLSKNEKIAFGAEOPRYOGSSSHINFR 365
QY 358 AVTPEDGLSKOLETPPSYVGAATVTPILGBOCLTPDAGPMRMRVNEPGETNRQHI 417
Db 366 TVSALE----- 371
QY 418 DELFTKTMHLVDYHPRADTWYADMEGTATDEDCYELGLVCGTAKAYVDDQLVY 477
Db 372 -----AVDGM-----ENTYAKG----- 384
QY 478 DNATQVPGDAFPGSATRETRGRINLVKNTYKFKIEGSAATYTLKGDYIPGHSILRV 537
Db 385 -----FSLNDEI----- 392
QY 538 GGCCKVIDDAEIE-KSVALAKEHDVILCAGLNADMETEGADRASMKLPGLVDOLADVA 596
Db 393 -----NTELEQCAVEAMNADKVVYFAGLPDSFESGFPKRMOLPOCIDLIDKLS 444
QY 597 AANPTVVVMOTGTPEBFWL-----DATPAVIOAMYGNETNSIADVVFQDYNPSGK 650
Db 445 EVNPNIVVVLHNGAPVEMPFANGDEDSNSVKAILEMYLSGQAAGBAVVRILLEGEVNPSSGK 504
QY 651 LSLSPFKRLQDNPAPLNFTEAGRTLYGEDVYVGRYYEFADKOVNFPFGHLSSTTAF 710
Db 505 LAETPPLRLDNPSTYLNFPGEADIVKYSGLIFVGRYKEMMELVLPFGHLSSTTEFY 564
QY 711 SNLSVSH---KDGK-LSVSLSVKNTGVSVAQAQVADLYKPLQAKINPVKELKFAVE 766
Db 565 SDIKISSYISDKKAPFYEMTYTNSGSRGDEEIIQIETPLPTVYI-KPIKELKFEKVE 623
QY 767 LQPGETKAVTIEQEKYVAAYFDEBDQWCVKGYEVIIVSDSA 811
Db 624 LKAGESKRVFRLDSSAFYYSKIDHD-WLSESGYINILIGKSSA 667

```

RESULT 5

```

ID 008331 PRELIMINARY: PRY: 754 AA.
AC 008331;
DT 01-JUL-1997 (Tremblrel. 04, Created)
DT 01-JUL-1997 (Tremblrel. 04, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Beta-glucosidase (EC 3.2.1.21).
GN BGLZ.
OS Clostridium stercorarium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxId=1510;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NCBI 11754;
RA Zverlov V., Adelsberger H., Bronnenmeier K., Staudenbauer W.;
RT "Nucleotide sequence of the bglz gene from Clostridium stercorarium."
RL Submitted (Apr-1997) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-NCBI 11754;
RA Bronnenmeier K., Staudenbauer W.;
RT "Purification and properties of an extracellular beta-glucosidase from
the cellulolytic thermophile Clostridium stercorarium."
RL Appl. Microbiol. Biotechnol. 28:380-386(1988).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-NCBI 11754;
RA Schwarz W., Jauris S., Koubka M., Bronnenmeier K., Staudenbauer W.;
RT "Cloning and expression of Clostridium stercorarium cellulase genes in
Escherichia coli."
RL Biotechnol. Lett. 11:461-466(1989).
DR EMBL: 294045; CAB08072.1;
DR InterPro: IPR002772; GH_3C.
DR InterPro: IPR001764; GH_3N.
DR Pfam: Pf00933; Glyco_hydro_3; 1.

```

```

DR Pfam: Pf01915; Glyco_hydro_3-C; 1.
DR PRINTS: PR00133; GHYDRASE3.
DR PROSITE: PS00775; GLYCOSYL_HYDROL_F3; 1.
KW Glycosidase; Hydrolase.
SO SEQUENCE 754 AA; 85207 MW; 12D6CEAB28C443EB CRC64;

```

```

Query Match 24.98; Score 1095.5; DB 2; Length 754;
Best Local Similarity 30.34; Pred. No. 1.le-04;
Matches 258; Conservative 137; Mismatches 247; Indels 209; Gaps 15;

```

```

QY 5 DVEALIKKTLAEVVDLAGIDFWHTKALPKHGVPSLRFDPGNGVY-----GTRK 55
Db 4 DIKILISOMTLLEKASLCSALDAWTLKVERLGIPSTWVSDBGHGLKRODRPTDEK 62
QY 56 FNGVPACFPQGSTSLGTFNQTLLLEAGKMKGAIKASAHVILPTINNQRSPPLGRG 115
Db 63 -TVPATCFPRAVLASWNRLEVEKYGALGECQAGIAVLGPETNKLKPLSGRNF 121
QY 116 ESTGEDPFLAGLAALIRGIQSTGVQATTIKHFLCNDQEDRRMAYOSIVTERALREIYAL 175
Db 122 EYFSEDEYLSSEMARSHIKGVOSRGVSTSLKHPAANQEHRRMSVDAVIDERTLREIYLA 181
QY 176 PEOIAVRDSQPGAFMTAYNGINGVSCSENPRIYLDGMLRKEMGMDGLMSDWGTYSSTA 235
Db 182 SEGAIVKAKAKWTIMCSYINRNGEYASENKFLLTDLVRNMGEGIVSDGAVNERVK 241
QY 236 VVAGLDLEMPGPPRFREGTELTKFVNSNGKPIIHVIDQAREVLQFVKCAASGVTENGPET 295
Db 242 LEALDLEMPSPFGIGOKIYEAVKKELEBEVLDRIVERILNLIK----- 288
QY 296 TVNPTPTAA-----LLRKVNGEIVLLKNNVNLPLSKKKTLVGNPAKQATYHG 347
Db 289 AYDRKKNAGCYDRANHKLAREARECMVLKNEKDTILPRKOGTIAVIEEFAPKRRYOG 348
QY 348 GGSALAAIYAAVFPFDLSQLETPPSYTGATTVTPILGBOCLTPDAGPMRMRVNE 407
Db 349 GGS----- 352
QY 408 PRGPNRQHIDELFTKTDMLVDYHPRADTWYADMEGTATDEDCYELGLVCGTA 467
Db 353 -----HVNPTIMDSPYEIKKSAQNNMDVIAQGIYI----- 384
QY 468 KAYVDDQLVVDNATKQVPGDAFPGSATRETRGRINLVKNTYKFKIEGSAATYTLKGD 527
Db 385 -----EKDEP----- 389
QY 528 IVPGHSILRVGGCKVIDDAEIEKSVALAKEHDVILCAGLNADMETEGADRASMKLP 587
Db 390 -----DEKLEBAKOTALKADVAAYFAGLPBHYBCEGDRTHMRMPS 432
QY 588 IDOLIAVAAANPTVVVMOTGTPEBFWLADATPAVIOAMYGNETNSIADVVFQDYNP 647
Db 433 HCTLIEVAETVNVVYVVLONGSPVEMPMWIDKKGLLEAVLGQAMGRLP--IFCSETP 490
QY 648 -SGKLSLSPFKRLQDNPAPLNFTEAGRTLYGEDVYVGRYYEFADKDVNFPFGHLSST 706
Db 491 IPKLAETFPKQLSDNPSYLNFR-ERDRVEXREGIFVGRYKKNMEPLFPGYGLSYT 549
QY 707 TFAFSNLSVSHKD-----GKLSVSLSVKNTGVSVAQAQVADLYKPLQAKINPVKELKF 762
Db 550 TFEYGDGKTSRKISKEISDNENVTAVSVKAKNGDMAGKEIVLYLVYDIETSVIDR--RTEEF 607
QY 763 AKYELQPGETKAVTIEQEKYVAAYFDEBDQWCVKGYEVIIVSDSAK 813
Db 608 EKVELDQGEKIV-VEELDKRAFAIYHRYKD-WHETGEFRDFIGNSSDVIYKDKIYK 665
QY 814 DGVALRGKFTV 824
Db 666 STYTIKKRMTV 676

```

RESULT 6

Q60038

ID 060038 PRELIMINARY: PRT: 720 AA.
 AC 060038;
 DT 01-NOV-1996 (TREMblrel, 01, Created)
 DT 01-NOV-1996 (TREMblrel, 01, last sequence update)
 DT 01-MAR-2002 (TREMblrel, 20, last annotation update)
 DE Beta-glucosidase (EC 3.2.1.21).
 GN BglB.
 OS Thermotoga neapolitana.
 OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
 NCBI_TaxID=2337;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=22706-MC24;
 RX MEDLINE=98048478; PubMed=9387232;
 RA Zverlov V., Volkov I., Vellakodvorskaya T., Schwarz M.;
 RT "Thermotoga neapolitana bglB gene, upstream of lamA, encodes a highly
 RT thermostable beta-glucosidase that is a laminaribiose.";
 RL Microbiol. 143:3537-3542(1997).
 DR EMBL: 277856; CAB01407.1;
 DR InterPro: IPR002772; GH_3C.
 DR InterPro: IPR001764; GH_3M.
 DR Pfam: PF00933; Glyco_hydro_3_1.
 DR Pfam: PF01915; Glyco_hydro_3_C_1.
 DR PRINTS: PR00133; GLHYDRLASE3.
 DR PROSITE: PS00775; GLYCOSYL_HYDROL_F3_1.
 DR Glycosidase; Hydrolase.
 KW
 SQ SEQUENCE 720 AA; 81370 MW; 4248AD2C75A28EF CRC64;

Query Match 22.4%; Score 982; DB 2; Length 720;
 Best Local Similarity 30.8%; Pred. No. 3.8e-57;
 Matches 265; Conservative 131; Mismatches 275; Indels 190; Gaps 26;

6 VEAIIKLTLEAKVDLAGIDF-----HTKALPKHGVSRLTDPNG 49
 4 VNEIISLTLEKSEKSCSG--GWTSGVWVSKSHGRCGETHPVPKRGALPAFVLADBPAG 61
 50 VR-----GTFENGVPACPCGTSIGSTFNQTLLEAKMKGEKIAKSAVILGPTIM 105
 62 LRINPTRENDENTYTTAPVPEIMLASTWNRLLEEVGKANGEEVREYGVVLLGPAANI 121
 106 QRSPLGRGPESTIGEDPFLAGIAAALIRIGIOSTGVQATIKHFLCNDQEDRRMAYOSIVT 165
 122 HRNPLGRNFEYSEDPVLSGEAASFVGVOSGVACIKHFVANNQETINMVDITVI 181
 166 ERALREIYALPFOIAVRDSOPGAFMTAVNGINGVSCSENPXTLDMGLKREMGMDGLINS 225
 182 ERALREIYALPFOIAVRDSOPGAFMTAVNGINGVSCSENPXTLDMGLKREMGMDGLINS 241
 226 WYGYSTTEAVVAGLDLMEPP-----REFGETLKEVNSNGKPEHIVIDQAREVLO 278
 242 WYAGDNVEQKAGNDLIMPKAVYQVTERDEIEIMALKKEGKLEEVLDCEVNNILK 301
 279 FVKKCAASGVTEGPP---ETTVNNTPEP---AALLKRVNEGIVILKNNVLLPSKKK 332
 302 -----VLVNAPESEKNRYSNKPDLEKHAVAAYEAGEGVALLKNE--EALPLSENNK 351
 333 TLIVGPNNAKQATVHGSGSALLRAVYAVTPEDGLSKOLETPPSYVAGATYTPVILGEQCL 392
 352 TALEGTGQIEIKGTSSTGTHPKATISILEGKER----- 387
 388 -----GLN---FDEELAKIYEDYIKR-----MRETEYKPR-RDSW----- 419
 433 ECHTETELDVVCGTAKAVVDQVLVADNATKQVPGDAFFGSATRETS--RINLVKNTYK 510
 440 -----GTTI---KPLSENFLESEKVENL-----AKKNDVAIVYISRLSGCYD 460
 511 FLIEFGAPTYTLKGDITVPGHSLRVGCKVVDQAEIEKSAVLAHEHOVITICAGLNA 570
 461 RK-----PVKGDFTLS-----DDETDLITKVS--REFHD----- 488
 571 DMETEGADPASKMLPGVLDOLADVAANPNTVVVWGTGP--EEMPMLDATPAVIOAMYG 629

DB 489 -----GKIYIVLLNGSPVEYVSMRDLVDGILVWQA 520
 630 GNETGNSIADVVEGDNYPGSKLSLSPKRLQDNPAFL---NRTFAGRTIGEDVYVGR 686
 521 GGEGRIVADVLGRINPSSGKLTPTTFRDYSVPSTFPEGRKDPKQVYEDDIYVGR 580
 687 YTEFADKDVNFPFGHISYTFPAFNSLSVSHKDKLSVLSYKNTGVSYPGAOVLVYKP 746
 581 YDFFVEPAVERGYGLSYTFEFSDLNVSFDGETLRVGYRIENLTGKRAKREVSQYIK- 639
 747 LQAKINRPVYKELGFAKVE--LQGETKAVTIEDEKRYAAAFDEPRDQWCEKGYEYI 805
 640 APKKIDKPPQELKAFKTRILNGBSEEVLEIPVRLASFNGEE---WVVEAGEYEV 696
 806 VSDSSAKQVALLKFTVE 826
 697 VGASSR--NIKLTGFSVGE 714

RESULT 7

ID 09KXN2 PRELIMINARY: PRT: 721 AA.

AC 09KXN2;
 DT 01-NOV-1999 (TREMblrel, 12, Created)
 DT 01-NOV-1999 (TREMblrel, 12, last sequence update)
 DT 01-MAR-2002 (TREMblrel, 20, last annotation update)
 DE Beta-glucosidase.
 GN TM0025.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
 NCBI_TaxID=2336;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MSB8 / DSM 3109;
 RX MEDLINE=99287316; PubMed=10360571;
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
 RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
 RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
 RT Evidence for lateral gene transfer between Archaea and Bacteria from
 RT genome sequence of Thermotoga maritima.*;
 RL Nature 399:323-329(1999).
 DR EMBL: AE001690; AM035119.1;
 DR TIGR: TM0025;
 DR InterPro: IPR002772; GH_3C.
 DR InterPro: IPR001764; GH_3M.
 DR Pfam: PF00933; Glyco_hydro_3_1.
 DR Pfam: PF01915; Glyco_hydro_3_C_1.
 DR PRINTS: PR00133; GLHYDRLASE3.
 DR PROSITE: PS00775; GLYCOSYL_HYDROL_F3_1.
 DR Complete proteome.
 KW
 SQ SEQUENCE 721 AA; 81073 MW; 31DB99D23B4CC963 CRC64;

Query Match 21.8%; Score 959; DB 16; Length 721;
 Best Local Similarity 30.4%; Pred. No. 1.3e-55;
 Matches 262; Conservative 130; Mismatches 280; Indels 190; Gaps 26;

6 VEAIIKLTLEAKVDLAGIDF-----HTKALPKHGVSRLTDPNG 50
 4 IDEIISLTLEKSKLVVGVGLPGLFNGPSRVAAGATHTHPVPRIGIFPAFVLADBPAG 63
 51 R-----GTFENGVPACPCGTSIGSTFNQTLLEAKMKGEKIAKSAVILGPTIM 106
 64 RINPTRENDENTYTTAPVPEIMLASTWNRLLEEVGKANGEEVREYGVVLLGPAANI 123
 107 RSLPGRGPESTIGEDPFLAGIAAALIRIGIOSTGVQATIKHFLCNDQEDRRMAYOSIVT 166
 124 RNPPLGRNFEYSEDPVLSGEAASFVGVOSGVACIKHFVANNQETINMVDITVI 183
 167 RALREIYALPFOIAVRDSOPGAFMTAVNGINGVSCSENPXTLDMGLKREMGMDGLINS 226

```

Db 184 RALRELYLKGFETIAVKKAPMTVMSAYNKLNGKYSQSENMELTKVLEBEMGFDFVSDW 243
OY 227 YGTSTTEAVVAGLDLEMPGP-----RFRGELIKFNYSNGKPFTHVIDORAREVLOF 279
Db 244 YAGDNPVEQLKAGNDMIFKAKAYOVNTRERDEIEIMEALKSEEVLEDECVRNLK- 302
OY 280 VAKCAASVTEGPE---TYVNNPTET---AALLRVGNIEGIVLKKNNVLPFSKKKT 333
Db 303 -----VLVNAPEFKCYRSNKPDLSEHAEVAYEAGAGSVLLEN-NOVLPFDEMTYH 353
OY 334 LIVGNPAQATYHGGGSAALRAYAVTPFDGLSKOLETPPSYVGAATVPILGSOCLT 393
Db 354 AVFGTGOLETKGSGSDTHPRITISLGEIK----- 387
OY 394 PDGAGMRVFNPEPPTPNRHOIDELEFTTDMHLVDYHPRKADTWYADMEGTYTADE 453
Db 388 -----RNMKFEDELASTYEYIKK-----MRETEKPR-TDSW----- 420
OY 454 DCTYELGVYCGTAKAVVDQVLVDNATKQVPGDAFFGSATREETG--RINLVKNTYKF 511
Db 421 -----GTVI---KPKLPENFLSEKEIK-----AAKKNDVAVVVISRISGEGYDR 462
OY 512 KIEFGAPTYTLKGDITVPGHSLRVGCGCKVIDDOAEIKSVYALAKE-HDOYITACGLNA 570
Db 463 K-----PVKGDFTLS-----DDELEIKTV--SKEFHDQ----- 489
OY 571 DWETEGADRASKMLPGVLDLIADVAANPNVYVYVQGTPEEM--PWLDAFPVAVIQAWYG 629
Db 490 -----GKVVVLLINISPIEVAWSMRLVDGILLVMOA 521
OY 630 GNETGSIADVYEGDYNPSGKLSLSPFKRLQDNPAFL--NRTAGRTILGEDYTYGVR 686
Db 522 GQEMGRIVADVLGKINPSGKLTTPPKDYSDVPSSTPPEGPKDPQVYVEEDITYGVR 581
OY 687 YTEFADKQNVNPRGHSLSTTFPAFNSLSVSHDKGLSVLSKNTGSPVGAQVQLVYKP 746
Db 582 YTFTEVPAVEFGSLSTYKEFYKDLKAIIDGELRLSYITITNGDRAGEVSGVYIK- 640
OY 747 LQAKINRPVKELKGAKE--LQPEKTAVTIEBQEKVVAVFPDERQKQVEKGYEVI 805
Db 641 APKGIKIDKPFQELKAFHKTKLNPSESEISLELPLDLASFSDEK--WVESGEYEV 697
OY 806 VDSSSAKGVALRGKFTY-GE 826
Db 698 VGASSR---DIRLRDIFLVEGE 716

RESULT 8
O51843
AC 051843 PRELIMINARY: PRT: 828 AA.
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Cellulobiose.
GN CBA.
OS Cellulomonas blazotea.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococciaceae; Cellulomonadaceae; Cellulomonas.
OX NGBL_taxid=1709;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-ATCC 486;
RC MEDLINE-98172737; PubMed-9511746;
RA Wong W.K.R., Ali A., Chan W.K., Ho V., Lee N.T.K.;
RT "The cloning, expression and characterization of a cellulobiose gene
RT encoding a secretory enzyme from Cellulomonas blazotea.";
RL Gene 207:79-86(1998).
DR EMBL: AF005277; AAC38196.1;
DR InterPro: IPR002772; GH_3C.
DR InterPro: IPR001764; GH_3N.
DR Pfam: PF00933; Glyco_hydro_3; 1.

```

```

DR Pfam: PF01915; Glyco_hydro_3; 1.
DR PRINTS: PR00133; GLHYDLASE3.
DR PROSITE: PS00775; GLYCOSYL_HYDROL_F3; 1.
SQ SEQUENCE 828 AA; 86180 MW; 3E8BDE46922030D CRC64;

Query Match
Best Local Similarity 31.2%; Pred. No. 1.8e-55;
Matches 262; Conservative 130; Mismatches 355; Indels 93; Gaps 17;

OY 4 IDEVALIKTLAKVLDLGLAGIDFEMHTKALPKHGVPSLRTDPNRYGKTFPNVPAAC 63
Db 7 LDPALVALASPLETKVALLTGATATFTLAPESIGLEVRISDPTGRLKFSGGRVAL 66
OY 64 FPGSTIGSTFNOITLLEBAKMMGKEAIAKSAHVILGPTINMORSPLGREGFSGEDPF 123
Db 67 FPNATLLASAMSEESTTEVGRLLAEALAQOHVVLGPTINLHRSVGLLEASSEDLL 126
OY 124 LAGGAALIRGIQSTGYOATIKHFLCNOQEDRRMNVOSITVERALREYALPFOIARD 183
Db 127 LTGRLLAAVYRGADLVGAGCLHLVANSETERNTMNSVVDPAFLRELYLPFEIAYDE 186
OY 184 SQGAFMTAVNGINGVSCSENPRTYLDMLRKEMGMDGLMSDYGSTTEAVVAGLDLE 243
Db 187 SDPWSVAAVNDVNGVPATEHHVYNEVLKEMGYTGLVMSMFEATRTAAPAAGGLDLV 246
OY 244 MPGRPRRGETLKFENSNKRPFIHVTDORAREVLOFVKCAASGVTEGPEPTVNNPTET 303
Db 247 MPGRGPMGDALVAARSGELDESVDHLRLRLVLAARVGLDLRDPDL--PAVDS 304
OY 304 AA---LIRKVNIEGIVLLENVNLPLSKKKTLLVGNPAQATYHGGGSAALRAYAVT 360
Db 305 AVREQLTRILAAAGMTVLNADDTLPIARGTRVALVGNHLETTIDMGGSATVPPQVS 364
OY 361 PREGLSKOLETPPSITVGAATVPILBQ--CLTPD-GAGMR-----WRVFENPFG 410
Db 365 VAEGLTALGADVVDVGEVTRPARPGFVVDPDGRGLHLTLAAGTYLDERHD 424
OY 411 TPNRHOIDELEFTTDMHLVDYHHPKA-----ADTYADMEGT 448
Db 425 APS-----TVMGGFDDPQAAVARFRARVAGBAGLEVGAIYGRMOTVAGGT 473
OY 449 YTADECTYELGLVCGTAKAYVDDQVLVDNATKQ-----VPGDAFFGSATREETGR 500
Db 474 -----ELAMTLATSGTGAEMELAPRTIQVHVHGSADVADATVVLRSSSTRSVT-- 522
OY 501 INLVKNTYKFKIEGSAFTYTLKGDITVPGHSLRVGCGKVIDDOAE---IKSVALK 557
Db 523 -----VGDADPCTDAGAAAEPLAGVGLGLVARRPAPEAEDVITRAAAAAA 568
OY 558 EHDYITACGLNADMETEGADRASKMLPGVLDLIADVAANPNVYVYVQGTPEEMPW 617
Db 569 QADAVVVVVGLTLEETTESYDKSTIALPGADALVRAVAAAARRTVVVVAATPVLMPL 628
OY 618 DATPAVIOAWYGNMTGNSIADVYFGDYNPSGKLSLSPFKRLQDNPAFLNRTAGRTLY 677
Db 629 DDVAVVIMAGLPGGEGHAAVAAALIGDQPTGRVLTTPPAAGCAAPAW--SVPPVGDLEY 687
OY 678 GEDVYVGRYEEFADK--DVNPFPGHGLSTYTPAFNSLSVSHDKGLSVLSVKNTGSP 735
Db 688 TEGFRFVYRGH--WADRAPAPAFMIGHGLGYATWEVADATLDDGAPAVTYVTITGART 746
OY 736 GAQVAVLYKPLQAKINRPVKELKGAFAVEIQPGTCAVTIEDEKVVYAVFDERDQW 795
Db 747 SREYVQVYTLTE---ASSDEVR-LVGMADATVDAGASARVVTADAR--MRRWDEAAG 801

RESULT 9
O87852
AC 087852 PRELIMINARY: PRT: 859 AA.
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)

```


DE Putative beta-glucosidase.
GN SC06597 OR SC0618.
OC Streptomyces coelicolor.
OC Bacteria: Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA MEDLINE=97000351; PubMed=8843436;
RX Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabbittush E., Rajandream M.A., Rutherford O., Rutter S., Taylor K.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2)."
RL Nature 417:141-147(2002).
RL EMBL: AL031013; CAI:19790.1; -
DR InterPro: IPR002772; GH_3C.
DR Pfam: PF00933; Glyco_hydro_3; 1.
DR Pfam: PF01915; Glyco_hydro_3; 1.
DR PRINTS: PR00133; GLHYDRLASE3
DR SEQUENCE 859 AA; 89455 MW; BD7EAFEA4FDD59F CRC64;
SO
Query Match 21.8%; Score 957; DB 16; Length 859;
Best Local Similarity 29.7%; Pred. No. 2,4e-55;
Matches 266; Conservative 131; Mismatches 354; Indels 146; Gaps 22;

DB DSPGAGADGAGAGPSSGAGLPGRPAHGAKSGPRPRRAGGRALARRAAGAVLL 361
QY KNNENVLLSKKK--KTLIVGNKQATYHGGSGAALRAYVTPFDLSKQL----- 369
DB ANK-DVLPDPDLPHCTAVVIGAHAKRTTGGSGAGVPRREVSVDGIRAEFGRRAYV 420
QY ETPPSYTGAVTTPVPIEGQCLTP-DGAPGMRMVFN-----EPPGT 411
DB HYPGRPDG---PAPLPDPDCTDPRSGLPVLLMLDADRELAYERRRGRLLEPLV 477
QY 412 ENRQHIDELFTKTDMLVDY---YHKAADTWADMGTTADEDCYVEGLVYCGRAK 468
DB PG-----AHVEIARLCPRTGSGW-----SLGVAGFGMS 508
QY 469 AYVDQIVY-----DNATKOYPGAFFGSATREETGRINLVKNTYKFEFGSAP 519
DB 509 LTTDGRTILEGDFPPSTDDPAVMHVPRAQYATA-----DLNAGR-----D 549
QY 520 TYTLKADITVPGHG--SLRVGCKVIDDQAEIENKVALAKEHDVYITAGLNDMEYGA 577
DB 550 TLLVARRELAPGTGRATVLAAPAPADVTASLAEAVRAAGADAADVAVVGTTEHESG 609
QY 578 DRASKLPGLVDQLADVAANPNTVMQGTPEEMQLATPAVIAWNGNETGNSI 637
DB 610 DRTDIALGATODALVRAAANAAPRTVAVNSGPELPEPREGAGAVLAMPFGDEGGGL 669
QY 638 ADVFPGDYNPSGKLSLSPFKRLQDNPAFLNFTAGRTLYGDEVYVGYRYEFADKYNF 697
DB 670 ADVLRHAPRPGRLPTTAVLADAPV-TRTPRDSGRIDYDGLHNGHMLRHRTAY 728
QY 698 PEGHGLSYTTFEAFSUNS---VSHKDGKLSVSVKNTGSPVGAQVLAQVY-KPIQAANIN 753
DB 729 WGHGIGYTTTWRHEELTLPVTRAGDGLVVRVRRVTGARAREVVOYVLIAP-ASALD 786
QY 754 RPKKELKGAHYELDQGEKATVIEQEKYVAAPDEEDDMCCKEYDEVIVSDSS 810
DB 787 RPARMLAGTVAARAREGTATVTRYPAR-ALRHMSVAEHAMRTAGPCRVLAGNSA 842

RESULT 10
ID 070021 PRELIMINARY; PRT; 808 AA.
AC 070021;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Beta-glucosidase.
GN Eryth.
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria: Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Pseudonocardiales; Pseudonocardiaceae;
OC Saccharopolyspora.
OX NCBI_TaxID=1836;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NRRL2338;
RX MEDLINE=98273631; PubMed=9613575;
RA Gaiser S., Boehm G.A., Douthett M., Raynal M.C., Dillon N.,
RA Cortes J., Leadlay P.F.;
RT "Analysis of Eryth and ErythII from the erythromycin biosynthetic gene
RT cluster in Saccharopolyspora erythraea.";
RL Mol. Gen. Genet. 258:78-88(1998).
DR EMBL: Y14327; CAA74702.1; -
DR InterPro: IPR002772; GH_3C.
DR InterPro: IPR001764; GH_3N.
DR Pfam: PF00933; Glyco_hydro_3; 1.
DR Pfam: PF01915; Glyco_hydro_3; 1.
DR PRINTS: PR00133; GLHYDRLASE3.
DR SEQUENCE 808 AA; 86677 MW; 0C619016CC00A751 CRC64;
SO
Query Match 21.7%; Score 954; DB 2; Length 808;
Best Local Similarity 30.8%; Pred. No. 3,4e-55;

Matches 260; Conservative 128; Mismatches 334; Indels 122; Gaps 21;

```

OY 6 VEAILKRLAEKVDLAGIDFMH-----TKALPKGVSLRTPDNVGRGK 54
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 37 VDELIGQLIDELKSLFV-----YWDYNEKPLAKMLPGVPLIGTFOIKRIGDSDPAGVTHHQ 92
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 55 FENGVPACFPCTSIGSTFNQTLLEAGKMGKKAIAKSAHVILGPTINMORSPILGGRG 114
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 93 -----PAIMPAFPAVALASAFDDRLAHEYGTVLGREGRAFEODIILIGPVNNIRVPOAGN 147
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 115 FESIGDPLLAGAALRGIOSTGVQATIHFLCNDQEDRRMAYOSIVTERALREIYA 174
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 148 FEFESDPLVLTARTAAQRIHSGGLMSAHYANTQEDRTIDVDQORTLRETEL 207
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 175 LPFOIIVRDSOGAFMTAINGVSCSENPYLLGMJLRKEMGMDGLMSDYGSTYTE 234
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 208 PGFEAAVA-AGATSVWCAIPKYNHACHORLLLEIKQMGFGVWASDWTATHA-TE 265
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 235 AVYAGIDLEMPGPP-----FRGELTKFVNSNGKPFTHVIDORAREVLOFYKCA 285
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 266 DIVAGIDQMGVEVEDGSLFRGKYLGEALKAIREGRIPESALDASVRILTQFEREGL 325
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 286 SGVTENGPEPTVNNTPETALLRKVNGEIVLLKNENVLPR--SKKKTLVGNNAQA 343
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 326 --LDEKPPRPBRDVAAGTRIAQEAESGAVLLRNEGGVLLDPAAGODIATIGSAQP 383
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 344 TVHGGSAALRAYAVTPEDLSKOLET--PESYVGAITYVTPILGOCITPDGAPMR 401
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 384 KYTGAGSYVEBDPANAPLIDITQVRVSGGARGVSGELGAP1-PTALQAPNVP6-- 440
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 402 WRFVNEPPTPKOHIDELFFTKTMHLVDYIHPRKADTWYADMEGYTAEDECTYELGL 461
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 441 -----EYTP-----PPSGGVYI--DGRILKVDADGLYRTAA 468
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 462 VVCGTAKAYVDQLVVDNATKQVPGDAFEGSATREETGRIMLVKNTYKFKIERG--SA 518
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 469 RIDG-----GNGSLQIDGAPIG-----VGDV-----FGPLTSV 497
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 519 PYYTLKGTIYFGHSLRVGCKVID-----QAEIKSVYALAKENHOVITICAGLN 569
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 498 PWWLTKGHEITOMTGAALVGGGSLDLDVLTWYPGHAGREFDAVAERARDSDVAVVA-- 554
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 570 ADMETEGADRASKLPGLDOLINDVANAANNTVVMOTGRPEEMPLDAPVIAQAMYG 629
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 555 YDDGELTADRTSLSLPGQODKLIDVAVSNNTVYVLTGSSVTMPMDKTRAVLDMWYR 614
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 630 GNETGNSIADVYFGDYNFSGKLSLSEFPKRLQDNPAF--LNFRTGAGRTLIGEDVYVGR 686
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 615 GQAGAEATTTALLFGDAEFGHRLGTFFPVSQERTIPVGGDPARFPGVDGVHSEGIFFSGYR 674
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 687 YTEFADKVNPPFGHGLSYTTPFANSLSVSHKDGKLSYSVKNTGVPQAQVQLYVKP 746
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 675 WYDRBGVPLPFPFGHLSYTFEERTDVPVETRGDLDTVTVRNTGORGSDVYQVYLPG 734
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 747 LQAAKINPVKELGFAVELPGETKAVTTEBOEKYVAAEFEDERQWCEKEDYEVY 806
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 735 SPQVPLDAPKQLAGYQVVELAPGETKRVYHVAER-ALOHMDAAGAKWLGKGRAVEI 793
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 807 SDS 810
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 794 GSS 797
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 11

09A7G7 PRELIMINARY; PRT; 762 AA.
 AC 09A7G7;
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DE 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE Beta-D-glucosidase.
 CC1756.
 OS Caulobacter crescentus.

OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
 OC Caulobacter.
 OX NCBI_TaxID=155892;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 19089 / CB15;
 RX MEDLINE=21173698; PubMed=11259647;
 RA Nierman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
 RA Eissen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
 RA Deboy R.T., Dodson R.J., Durkin A.S., Gwin M.L., Haft D.H.,
 RA Kolonay J.F., Smit J., Craven M.B., Kouri H., Shetty J., Berry K.,
 RA Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.,
 RT "Complete genome sequence of *Caulobacter crescentus*."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 DR EMBL; AE005849; AKK23732.1;
 DR TIGR; CC1756;
 DR InterPro; IPR002772; GH_3C.
 DR InterPro; IPR001764; GH_3N.
 DR InterPro; IPR002016; Peroxidase.
 DR Pfam; PF00933; Glyco_hydro_3_1.
 DR Pfam; PF01915; Glyco_hydro_3_1.
 DR PRINTS; PR00133; GLHYDLASE3.
 DR PROSITE; PS00436; PEROXIDASE_2; UNKNOWN_1.
 KW Complete proteome.
 SQ SEQUENCE 762 AA; 81011 MW; CAICCT6B0D7E0ZDB CRC64;

Query Match 21.6%; Score 946.5; DB 16; Length 762;
 Best Local Similarity 30.5%; Pred. No. 9, 9e-55;
 Matches 257; Conservative 115; Mismatches 273; Indels 199; Gaps 20;

```

OY 4 IDEVALKRLTLEKVDLAGI-----DEWHTKLRPKGVSLRFTGPN 49
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 55 VRADALVKQMTIDELKTYLHGLFPPAKRAPMTMISAGVPEVPLNPTLRPSASIG 114
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 50 VRG-TKFGVCPACFPCTSIGSTFNQTLLEAGKMGKKAIAKSAHVILGPTINMORS 108
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 115 VANOVEQRKGDVATLPSGLALSTFEPKLAFGAMIGAEARAKFPNVLAGVNLTRD 174
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 109 PLGRGFEISGEDEPLAGLAALINGISTQVQATIKHLCNDQDRRMVOSIYTERA 168
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 175 PMAGRNFEYGEDEPLAGEVAAQIKGVOSNKTIVSTIKHLMALQGTGRYMDAQDEAD 234
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 169 LREIYALPQIAVRDSQGAFTAYNGVSCSENPYLLGMJLRKEMGMDGLMSDYG 228
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 235 LRSDDLAFQIALEKSNPSPVCAKYNKVGDMACENDPLNKLKDKMYPGWNHSDWA 294
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 229 TYSTTEAVAGLLEMPGPP-----FRGELTKFVNSNGKPFTHVIDORAREVLOFYKCA 284
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 295 VHSIVKAAALGLD-QOSGELDTQIFGEBDKAAVAKGEVSQARVDVNRILHGV---I 350
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 285 ASGVTEGPEPTVNN--NTPETALLRKVNGEIVLLKNENVLPRSK-KKTLVGNPAK 341
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 351 SSGLMNPTTSQPLIDYAHAAVAVOTVAERGSVLLKNRGDLPLKSKAKKTVLGAHND 410
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 342 QATYHGGSAALRAYAVTPEDLSKOLET--PESYVGAITYVTPILGOCITPDGAPMR 401
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 411 VGVISGGSSQVR-----SVGAPVLEIPKGGPAAS----- 441
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 402 WRFVNEPPTPKOHIDELFFTKTMDHLVDYIHPRKADTWYADMEGYTAEDECTYELGL 461
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 442 -----EYTP-----PPSGGVYI--DGRILKVDADGLYRTAA 468
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 462 VVCGTAKAYVDQLVVDNATKQVPGDAFEGSATREETGRIMLVKNTYKFKIERG--SA 518
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 450 -----SSPLQ 454
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 522 TLKGDITVPGHSLRVGCKVIDDQAEIKSVYALAKENHOVITICAGLNADMETEGADRAS 581
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 455 AIKA-----ANSAEV---TYVDGKDPV-AAAAAAKADADVAIVFAM---HMQTEADAPS 502
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 582 MKLPGLDOLINDVANAANNTVVMOTGRPEEMPLDAPVIAQAMYGNETGNSIADV 641
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```



```

Db      503  IALPENODALIEAVSANKNAVVLEETGCVLMPWLCKGVAQLQWPGQGGQALIALL 562
QY      642  FGQVNPBGKLSLEPPKRLQDN-----AFLNFTREAGR-----TLXGEDVY 682
Db      563  FGEVNEPBGRLAMTFPPSEDOAPRASAAPRAEQALIDARAGOKAPKIGFVPRVEGMA 622
QY      683  VGYRYEFAADKDVNFPGHGLSYTTFASFNLVSYSKDKLSVLSYKNTGSPGAQVQOL 742
Db      623  VGYWFMQAEKRRLPLYPGGLSYTSFGYKMLKVEDGDG-LKVSFDVTNKGKAGADTQDL 681
QY      743  YVKPLQAKINRPKELKAGAKYELQPGETKATIEOEOKYVAAYPEDEBDQMCYEKPY 802
Db      682  VYTSQO-----RKAMLRLAGQKVDLAPGETKRVTLVNERILLADY-DTAKPGMTIAGTY 736
QY      803  EVIV 806
Db      737  PLVY 740

```

```

RESULT 12
O9ZGH5  PRELIMINARY; PRT; 809 AA.
ID O9ZGH5
AC O9ZGH5
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Beta-glucosidase.
GN DESR.
OS Streptomyces venezuelae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=54571;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC15439;
RC MEDLINE-98445333; PubMed-9770448;
RA Xue Y., Zhao L., Liu H.W., Sherman D.H.;
RT "A gene cluster for macrolide antibiotic biosynthesis in streptomyces
RT venezuelae: architecture of metabolic diversity."
RL Proc. Natl. Acad. Sci. U.S.A. 95:12111-12116(1998).
DR EMBL: AF079762; AAC68679.1;
DR InterPro: IPR002772; GH_3C.
DR InterPro: IPR001764; GH_3N.
DR Pfam: PF00933; Glyco_hydro.3; 1.
DR Pfam: PF01913; Glyco_hydro.3_C; 1.
DR PRINTS: PR00133; GLHYDRASE3.
SQ SEQUENCE 809 AA; 84686 MW; DE15489926BE37A8 CRC64;

```

```

Query Match 21.5%; Score 944.5; DB 2; Length 809;
Best Local Similarity 32.2%; Pred. No. 1.5e-54;
Matches 275; Conservative 112; Mismatches 325; Indels 143; Gaps 26;

```

```

QY 9 ILKLTALAEKVDLLAGIDFW-----HTKALPKHGVSRLRTDGPNGVGRKTFENG 58
Db 56 LVQVQMLDEKISFV-----FMALDPRQNYGLPGVPRIGLDELRAADPNRIR---LVG 107
QY 59 VPACPCPGSTGSTNQTLLLEAGKMGKEALAKSAHYIIGPTINMORSPGLGCFESI 118
Db 108 QVTAAPAPVALASTFDDTMAISYGVKVRDGRBALNODMVLGPMANNIRVHGNGNYTF 167
QY 119 GEDPFLAGLAALIRIGISTQVATIKHFLCNDQEDRRMAYQSIYTERALREIYALFQ 178
Db 168 SEDPVSRSRAVAQIKIGOGAGIMTAKHFAANNOENNFVNAVNDQTLREIFPAFE 227
QY 179 IAVRDSQPGA--FMTAYNGINGVSCENPKYIDGMIRKMGMDGLMSWGTSTYTA 236
Db 228 AS---SKAGAASPMCAVNGNLKRPSCGNDELNNVLTQMGQGVMSMLAT--FGTDAI 283
QY 237 VAGIDLEFM-----PCGP-RFRGETLKFNYSNCK--PFIHYIDQARAVLYQFYK 282
Db 284 TKGIDQEMGVLELPGDVKSGPSPAPKPFGLKLTAVLNGTVEAATVRSABRIYQMEKF 343

```

```

QY 283 CAASGVTENGPETTNNTEETALLKRVNGESIVILKNNENNVPLS--KKKTLIVGNA 340
Db 344 GILLATPAPRPE---RDKAGAGVSRKVAENGAIVLLRNKGALPLIAGAGSIAVIGPTA 400
QY 341 KQATYHGGGSAALRAVYATPPDGLSKQLEFPSTYVGAATVPPILGEOCLTPPGAAGM 400
Db 401 VDPKVTGLGSAHVHVPDSAAAPLDTIKAR-----AGAGATVYETGETGTGTQIPAGN 452
QY 401 RMRYENEPGTPNQRHIDELFTKIDMHLVDYHPRPADTWYADMEGYTAEDCTYELG 460
Db 453 LSPAFNQG-----HOLE--PGKAGALY---DGLTLPADDEYRIA 487
QY 461 LVWGTAKAVYDDQLVVDNATKQVPGDAFFGSAATRETRINLVGNNTKFFIEGSA 520
Db 488 VRATG---GYATVQGL---GSHTEAGQVYKYS---SPLKLTGK-TKLL----- 527
QY 521 YTKGDTIVPHGHSRLRYGCGKVIDDOAEIKSVALKAEHDOVITICAGLADNETEGADRA 580
Db 528 -TISGFAMASATPLSLGAWTPPAADATIKAVESARKATAVFA---YDGTGEGVDRP 583
QY 581 SMKLPVLDOLIALDVAANPNTVVVMTGTPPEMPWIDATPAVIOAWYGNENIGNSIADY 640
Db 584 NLSLEPGTDKLSAVADANPNTIVLVNTGSSVLMPLSKTRAVLDMMWYPC2AGCAETAL 643
QY 641 VEGDNPBGKLSLEPPKRLQDNPAFLNFTREAG-----RTLYGEDVYGYRYEE 690
Db 644 LYGDVNPBGKLTQSF-----PAENQHAVAGDPTSYPCVDMQYTRREGIHGYIMFK 696
QY 691 ADKDVNPFPGHGLSYTTFASFNLVSF-H-KDGLKLSVLSVNTGSSVPGAOVAYVYKPLQ 749
Db 697 ENVKLFFPGHGLSTSTQCAPTVVTRSTGGLVTVYVTRNSGRAGOEYVQAYLGASPN 756
QY 750 AKINRPVELKGFALVELQPGETKAVTIEOEKYYAAYPDEBRQMCVEKGVDEVYSDS 809
Db 757 VTAQAKKKLVGYTVKVSIAAGEAKTVTVN-----DRRQLOIGS 795
QY 810 SAKAGVALRGKFTY 824
Db 796 SSAD---LRGSATV 806

```

```

RESULT 13
O9P3Y0  PRELIMINARY; PRT; 781 AA.
ID O9P3Y0
AC O9P3Y0
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Exo-1,4-beta-glucosidase (EC 3.2.1.21).
GN Prevotella albensis.
OS Bacteria; CFB group; Bacteroidetes; Bacteroidales; Prevotellaceae;
OC Prevotella.
OX NCBI_TaxID=77768;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-M384;
RA Walker N.D., McEwan N.R., Wallace R.J.;
RT "exo-1,4-beta-glucosidase gene from Prevotella albensis."
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AJ276021; CAC07184.1;
DR InterPro: IPR002772; GH_3C.
DR InterPro: IPR001764; GH_3N.
DR Pfam: PF00933; Glyco_hydro.3; 1.
DR Pfam: PF01915; Glyco_hydro.3_C; 1.
DR PRINTS: PR00133; GLHYDRASE3.
SQ SEQUENCE 781 AA; 85866 MW; 7F025625B8735CB5 CRC64;

```

```

Query Match 21.4%; Score 941; DB 2; Length 781;
Best Local Similarity 29.6%; Pred. No. 2.4e-54;
Matches 257; Conservative 123; Mismatches 289; Indels 200; Gaps 24;
QY 5 DVEAILKLTALAEKVDLLAGI---DF-----WHTKALPKHGVSRLRF 43

```

```

Db 31 NIDVIAANTLEKKAOLLVGVGHODVSGTLMGSHSLVAGAACQTEISRLGIPATVY 90
OY 44 TDGPNCVR-----GTR--FENGVPACPCSTSLGTFENOTLLEAGKMKKKAIAKSA 95
Db 91 ADGPACVHINPRTPGTNOJF-----ATGPRTCTCLASTWINDLYHVAKAIGNETLXGI 146
OY 96 HVLGFTIMQSSPLGGRFESIGEDPFLAGAAALIRGIQSTGVQATIKHFLCNDQED 155
Db 147 DVLGGMNLMHSPCLGGRFESYSDPVTYGLIGSAMKGIOSOGVAGSAFHAFNANSES 206
OY 156 RRMVOSIYTERALREIYALPROIAYRDSOGAEWTAANGINGVSCSNRPYLDGMRLKE 215
Db 207 DTRFRDERISQALREIYALKEIYVROSKPMTLMSTNKTNGTYTQSGKDLNLIIRKD 266
OY 216 WGDGLIMSDWYGYTS-----TTEAVVAGLDEMGPPRFNGETLKFVNSNGKPFIVHYDQ 271
Db 267 WGYOGIVMTDWIGERADLPVETE--VEAGNDPMFGND--RAKHIVKAVKAGRLDIKIVAR 324
OY 272 RAREVLOFVKCAASGVTEENGPEYVNTPET-----AALIKVNEGIVLLKNNENVLPLS 328
Db 325 NIKNNLEIYTLK-----TPRTKTKYTNQPOLKAHAQITTRQASTEGVLLKNDNNVLPVK 378
OY 329 KKKKTLIVGPNKQATYHGGSSAALRAYAVTPEDGLSKOLETPPSYTGATYTPPIIG 388
Db 379 NKKKVALGVSNDYDLGSLGSGCVNPPYVDWVHGLO-----NAGIATF----- 423
OY 389 EGCCLTPDAPGMRKRVFNEPPTGTRKOHIDELFTTKTDMHLVDYHPRKADTYADMEGT 448
Db 424 -----KOLTEIYENVYKAKAKIADKRNEMFLOD----- 454
OY 449 YTADECTYEELGVCGTAKAYVDDOLVVDNATKQVDFAFGSAFTEEGRLINLYKNGT 508
Db 455 -----GQPK--LDEIETQRCVHEVGDA-----DAIITIKR--- 485
OY 509 YKFKIEPSAFTYTLKGDYTPGHGSLRVGCGKVIDDOAEIEKSVALKHEHDVITICAGL 568
Db 486 -----QAQEGMDRSIEGEFNLTDH-----EKAMISR----- 511
OY 569 NADWETEGADRASMKLPGVLDOLADVAAN--PMTVYVVOGTPTPEMPLDITPAVITOA 626
Db 512 -----VSDVFAHNNKFPVYIINSGSVMTETASMRDVAIIYA 548
OY 627 WYGNETGNSIADYVFGDYNPSCGLSLSP-----KRLQDNPAFLNFRTE----- 671
Db 549 WQPEEGEGNSYADVLIGKANPSGHLTSTWPIASADYPSRKPNFQQAAYINLSDKILXNMA 608
OY 672 --AGRTLYGEDVYVGYRYEFAKDNVFPBGHLSITTAFAISMLSVSHKDGKLSVLSYK 729
Db 609 KGVNTTNEEDYVGYRYEFTFNKKVAYPFGYGLSYTTEFGKPSVSLNGDKITVTVSVK 668
OY 730 NTGVSQAQVAYOLVYKPLQAKINRPVKELKFAKYE--LQPGETKAVTIEBQEKVAAVF 788
Db 669 NIGVAKQVAYVYVAPKGA--YERPSCELKAFATKMLKPOSELTLMKIATKRLAS-F 726
OY 789 DEEDDQCVKEDYEVIVSDSSAAKDGA 817
Db 727 DEANSQMKYDAGKYEEFCVGDWVESIKGTA 755

```

RESULT 14

```

088C29 PRELIMINARY; PRT; 793 AA.
AC 088C29;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Putative beta-glucosidase (EC 3.2.1.21).
GN EGLB OR YP02803.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
NCBI_TaxID=632;

```

```

RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=CO-92 / BIOVAR ORIENTALIS;
RX MEDLINE=21470413; PubMed=11586360;
RA Parikh J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Partridge M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Felwell T., Hamlin N., Holtroyd S., Jagers K., Kariyasev A.V.,
RA Leather S., Koutle S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
DR EMBL: AJ141453; CAC93037.1;
DR InterPro: IPR002772; GH_3C.
DR InterPro: IPR001764; GH_3N.
DR Pfam: PF00933; Glyco_hydro_3; 1.
DR Pfam: PF01915; Glyco_hydro_3_C; 1.
DR PRINTS: PR00133; GLHYDRASE3.
DR PROSITE: PS00775; GLYCOSYL-HYDROL_F3; 1.
KW Hydrolase; Glycosidase; Hypothetical protein; Complete proteome.
SQ SEQUENCE 793 AA; 86853 MW; 6C23D6093998955 CRC64;

```

```

Query Match 21.4% Score 939.5; DB 16; Length 793;
Best Local Similarity 28.7%; Pred. No. 3.1e-54;
Matches 242; Conservative 125; Mismatches 266; Indels 211; Gaps 16;

```

```

OY 6 VEAIIKRLTLEKVDLLAGIDF-----WITRALPKHGVPSLRFIDGPNQVR----- 51
Db 31 IPSVIAANTLEKKAFAVSGTGMNAEKVAVAGASTLAIPRLGIPQIVFADPGVYRLAG 90
OY 52 ---GTRFENGVPACPCSTSLGTFENOTLLEAGKMKKKAIAKSAHILGPTIMQMS 108
Db 91 PTGGEKRP-----ATGFPVSAAMATMDPALIKRVGALIDGMAQYVVDLILGPAIGH 146
OY 109 PLGGRFESIGEDPFLAGAAALIRGIQSTGVQATIKHFLCNDQEDRPMVQSVITERA 168
Db 147 PLNGRNEFTEDPLNAINTASTYIDGMQTEGVAALKHFAANNQDRETRQIYNEIISDRA 206
OY 169 LREIYALPFOIAYRDSOGAFMTAYNGINGVSCSNRPYLDGLRKREMGDLIMSDWYG 228
Db 207 LHEIYFPGFEVAMEKQAPMAVMSYPSINTPSSONPMLIKVLRQOKKFDGFVMSDWYG 266
OY 229 TYSTTEVAVAGLDEMG-----PPRFGETLKFVNSNGKPFIVIDQAREVL 277
Db 267 VADVPALKGNDLNMGEGRTPDLSFLTPNTDPKVVYLAALKSELTQDQIDENIRIL 326
OY 278 QFVKAASGVTEENGPEYVNTPETALLIRKVGNGSYLLKNNENVLPI,SKKKKTLIVG 337
Db 327 NVYIK-----THRFKNGDAAMPEPEVDHSALAKQIADSMVLLKNDHATLP,IKAAFPVAAFG 383
OY 338 PNAKQATYHGGSSAALRAYAVTPEDGLSKOLETPPSYTGAVTYTPPIIGROCLTPDGA 397
Db 384 QNIDNFEVVGSGSAEYV---IDP-----KRLVTLPOGLKXNAGIALINQIGDAL----- 429
OY 398 PGRKRVFNEPPTGTRKOHIDELFTTKTDMHLVDYHPRKADTYADMEGTYTADECTY 457
Db 430 -----DEPVNP----- 436
OY 458 ELGLVVGCTAKAYVDDOLVVDNATKQVDFAFGSAFTEEGRLINLYKNGTIRKIEFGS 517
Db 437 -----ALTKA----- 441
OY 518 APTYTLKGDYTPGHGSLRVGCGKVIDDOAEIEKSVALKHEHDVITICAGLNADMEGA 577
Db 442 -----AQSDDLALISIGSS---TEGA 460
OY 578 DRASMKLPVLDOLADVAAN--NPTVYVVOGTGPEER--PVLDAFPAVIAQWYNGNE 632
Db 461 DRTSAMHMDPEVAMIKRYASAFOQNKRVYVLLNIGSIPIEMTSWEQYADAAILTWGCEQ 520
OY 633 TGNSTADVVFQDYNPSCGLSLSPKRLQDNPAFLNFRTEAGRTLYGDDYVGVRYRYEAD 692

```

Db 521 AGNAVADILLGKVPNSCKLPLTPFKRLDESDSPSGNPNQNAKTVYIGGIVGYRFPDKR 580
 QY 693 KDVNPPGCHLSYTFPAFNSLVS-----HKDGKLSVLSYKNTGSGVPAQVQLVYKP 746
 Db 561 IAPMPFGYGLSTVYNGKISPEKPVFNIDTENSIEVSLPRTSGIDTKEVQLVHD 640
 QY 747 LQAARKINPVELKGFALVLOPGETKAVTIEOEKYYAAVFEDEPDQMCVEKGDYEVIV 806
 Db 641 -NASHDRPEBELKAFERKVSLSAGEKRYAF-KIDKRAFSTYEDDKNNWVLEPOLFTLRI 698
 QY 807 SDSS 810
 Db 699 GRSS 702

RESULT 15

Q9RH03 PRELIMINARY; PRT; 732 AA.
 ID Q9RH03
 AC Q9RH03;
 DT 01-MAY-2000 (Tremblrel, 13, Created)
 DT 01-MAY-2000 (Tremblrel, 13, Last sequence update)
 DT 01-MAR-2002 (Tremblrel, 20, Last annotation update)
 DE Beta-glucosidase precursor.
 GN SALA.
 OS Azospirillum irakense.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillaceae;
 OC Azospirillum.
 OX NCBI_TaxID=34011;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STAIN-KR01;
 RA Faure D., Desair J., Keljers V., Proost P., Henriessat B.,
 RA Vanderleyden J.;
 RT "The sala and salb genes encode two beta-glucosidases in Azospirillum
 irakense.";
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF090429; AAF21798.1;
 DR InterPro; IPR002772; GH_3C.
 DR InterPro; IPR001764; GH_3N.
 DR Pfam; PF00933; Glyco_hydro_3; 1.
 DR Pfam; PF00933; Glyco_hydro_3; 1.
 DR PRINTS; PR00133; GLHYDLASE3.
 KW Signal.
 FT SIGNAL.
 SO SEQUENCE 1 26 POTENTIAL.
 F792A947923F4E1E CRC64;

Query Match 20.9%; Score 919.5; DB 2; Length 732;
 Best local similarity 30.7%; Pred. No. 6e-53;

Matches 260; Conservative 106; Mismatches 275; Indels 205; Gaps 23;

QY 12 KLTLAEKVDLAG-----IDFWHTKALPKHGVPSLRFTDGPNGVR 51
 Db 49 ELTIDERSLHGMPLPFPSPPTIPGSPSLVPIF--PCVPRLGIPALKETDASLGV- 104
 QY 52 GTRFEF--GVPAACFPQSTSLGSTFNQTLLEAGKMKGEAKSAHVILGPTINQRS 108
 Db 105 -TNPNNVAPGDTAALPSGALASTFNPRLKSLDGGAAIAKKAASKGFNVILAGANLARD 163
 QY 109 PLGGRGFESIEDPPLAGLAALIRGIQSTGVQATIKHPLCNDQEDRRMVOISIVTERA 168
 Db 164 PRGNRNFYLDGDPDLGILAGESIRGIOSONIIISTYKHPSLNGOETNRHMGNSVIDEAA 223
 QY 169 LREIYALPFOIAVDSOGAFMTAYNGINGVSCSENPKYLDGMLCKEMGMDGLTMSDYG 228
 Db 224 HRESDLAFOIAIERGDPGSVMCAIYNVNGAYSCGNHLLKYLKGDWGYKGMVMSDW-G 282
 QY 229 TYSTTEAVVAGLDEMPGPR---FRGETLKFNVSNKCPFIHVIDQAREVLQFVKCA 284
 Db 283 AVPATDFALKGLD-QOSQQLDEKIFWGDLLKEAAAAGTIPAERLSDMSRRIL---RSMF 338
 QY 285 ASGYTENGPETVYNTPTALLKRVNGEIVLLKNNENVIPLSK-KKTLIVGPNAKQA 343
 Db 339 AAGFFDGKRGKPVYDLDAHAIAIAKOVADGIVLLANDKGLPLAAGSOKIAVIGGFADOG 398

QY 344 TYHGGSAALRAYAVTAPEDGLSKOLETPPSYVGAVTTPPLIGBOCLTPGAPGMRMR 403
 Db 399 VLSGAGSSQYT-----SVGNPPIVPIPGSGML-----AATLRQA 433
 QY 404 VFNEPPGTPNROHIDELFFKTDMLVDYHHPKAADTWADMGTYTADEDCYELGLVY 463
 Db 434 YHNSPLKALKERL-----PNATIRF--NDGRYSA----- 461
 QY 464 CGTAKAYVDQLVYDNATKOVPGDAFFGSAITRETRINLVKNTYKFKIERGSAFTYTL 523
 Db 462 ----- 461
 QY 524 KGDTIVPGHSLAVGGCKVIDQAIEKSVATAKKEHDYIICAGLNADMETEGADRASK 583
 Db 462 -----AAALAKQSDIYILFAN--QWMSBGMADYDLK 491
 QY 584 LPGVLDQLADVAANPNVTVVWQGTPEBMPWLDATPAVIOAWYSGNETGNSIADVEG 643
 Db 492 LPGQDALIETVAEAMPNAVIVLQGTGPVLMPWKDKVGAIVSAWYSGQGGGALADILVG 551
 QY 644 DYNPSCGLSLSPFKRLQ-----DNPAFLNRTBAGRTLYGEDYVYGRYTERFADKVN-- 696
 Db 552 KTNPSGRLPSTFPASADQYPHPEVPGMNLPEKCOFPDVEEESDVGVR--REAAKGMKPL 609
 QY 697 FPPGHGHSYTFPAFNSLVSXKDKG-LSYLSYKNTGSGVPAQVQLVYKPLQAAKINRP 755
 Db 610 FPPGHGHSYTFPAFNSLVSXKDKG-LSYLSYKNTGSGVPAQVQLVYKPLQAAKINRP 755
 QY 756 VKELGFAKVELOPGETKAVTIEOEKYYAAVFEDEPDQMCVEKGDYEVIVSDSSAAKDG 815
 Db 667 ---LIGFEKIDLPKGRRRVYTIKADRLAR--FDEGQHWRIDGSDYDVVVGSAATM--- 719
 QY 816 VALRGK 821
 Db 720 TVLSGR 725

Search completed: April 26, 2003, 13:09:46
 Job time : 52 secs

